

Running head: VERBAL PROTOCOLS AND THE SELECTION TASK

Think-Aloud Protocols and the Selection Task: Evidence for Relevance Effects and
Rationalization Processes

Erica J. Lucas

Nottingham-Trent University, UK

Linden J. Ball*

Lancaster University, UK

*Corresponding author:

Linden J. Ball, Psychology Department, Lancaster University, Lancaster, LA1 4YF, UK

Tel: 01524 593470

Fax: 01524 593744

Email: L.Ball@lancaster.ac.uk

Abstract

Two experiments are reported that employed think-aloud methods to test predictions concerning relevance effects and rationalization processes derivable from Evans' (1996) heuristic-analytic theory of the selection task. Evans' account proposes that card selections are triggered by relevance-determining heuristics, with analytic processing serving merely to rationalize heuristically-cued decisions. As such, selected cards should be associated with more references to both their facing and their hidden sides than rejected cards, which are not subjected to analytic rationalization. Experiment 1 used a standard selection-task paradigm, with negative components permuted through abstract conditional rules. Support was found for all heuristic-analytic predictions. This evidence was shown to be robust in Experiment 2, where "select-don't select" decisions were enforced for all cards. Both experiments also clarify the role played by secondary heuristics in cueing the consideration of hidden card values during rationalization. We suggest that whilst Evans' heuristic-analytic model and Oaksford and Chater's (e.g., 2003) optimal data selection model can provide compelling accounts of our protocol findings, the mental models theory fares less well as an explanation of our full dataset.

Acknowledgements

This research was conducted whilst Erica Lucas was at the University of Derby, and we acknowledge the University for its financial support. We are grateful to Susanne Hempel, Andrew Morley, and Nicki Morley for valuable discussions relating to the reported studies.

Think-Aloud Protocols and the Selection Task: Evidence for Relevance Effects and Rationalization Processes

The nature of human reasoning has long been of interest to cognitive psychologists. One task in particular continues to attract considerable research attention, namely, the abstract version of Wason's four-card selection task (e.g., Wason, 1966). Although this task appears straightforward, few people respond with logically correct selections. In its most common format, participants are presented with an array of four cards, and are told that each card has a letter on one side and a number on the other side (only the facing sides are visible). Participants are then given a conditional rule that they are told applies to the cards, and their task is to decide which cards need to be turned over in order to determine whether the rule is true or false. For example, the rule might be 'If there is an A on one side of the card then there is a 3 on the other side of the card', and the cards might be 'A', 'J', '3' and '7'. These are referred to as the p, not-p, q and not-q cards, respectively (p and not-p are antecedent cases and q and not-q are consequent cases). The logically correct response for a conditional reading of the rule is to turn the A (p) and the 7 (not-q) cards, as these could potentially provide a letter-number combination that would show the rule to be false. For a biconditional reading of the rule it would be necessary to select all four cards to check for a falsifying letter-number combination. Most people, however, select either A (p) alone, or A (p) and 3 (q).

To explain these choices, Wason (e.g., Wason & Johnson-Laird, 1972) suggested that people were displaying a verification bias, whereby they were trying to prove the rule true (by finding a card with the A and 3 combination), rather than trying to prove it false, as logic necessitates. However, Evans and Lynch (1973) demonstrated that when negative components are introduced into the conditional people's selections indicate a systematic matching bias rather than a verification bias, in that they simply seem to

choose cards that are named in the given rule. Matching bias is a robust phenomenon in selection tasks employing abstract rules incorporating the following connectives: if p then q, q if p, p only if q, and there is not both p and q (e.g., Evans, Clibbens, & Rood, 1996; Evans, Legrenzi, & Girotto, 1999). Recently, too, Roberts (2002) has demonstrated similar levels of matching to that seen with conditionals in selection tasks involving categorical rules of the form all p have q. Finally, matching has been observed in paradigms other than the selection task such as truth table tasks -- where people have either to identify or construct instances that verify, falsify, or are irrelevant to a given conditional or categorical rule (e.g., Evans, 1998b; Evans et al., 1999).

Despite the apparent generality of the matching phenomenon there is some remaining contention as to whether it extends beyond conditional and categorical rule forms to rules such as inclusive disjunctions (either not p or q, or both) and exclusive disjunctions (either not p or q, but not both). Evidence from truth table tasks suggests such generality (Evans et al., 1999; Evans & Newstead, 1980), whilst evidence from disjunctive selection tasks is highly inconsistent: Evans et al. (1999) observed weaker matching effects than arise with other rule forms; Krauth and Berchtold-Neumann (1988) found no matching with inclusive disjunctions but an effect with exclusive disjunctions; and Roberts (2002) demonstrated inverted matching (i.e. reliably fewer matching than mismatching selections). We concur with Roberts' (2002, p. 96) view that the safest inference from current findings is probably that there is no matching effect on disjunctive selections tasks, and that the idea that matching is a general linguistic phenomenon (e.g., Evans', 1998b) is uncertain in as much as it may only occur for conditional and categorical statements that have an inherent directionality.

Theoretical Accounts of Matching Bias in the Selection Task

Three reasoning theories of major contemporary importance have been applied to selection tasks involving abstract conditionals in an attempt to explain people's choices, including the dominant matching pattern. Evans' heuristic-analytic (henceforth H-A) theory (e.g., Evans, 1989, 1996) proposes that reasoning involves two distinct processing stages. First, implicit, pre-conscious, heuristics determine which aspects of a given task are of psychological relevance, thereby enabling attention to be selectively focused on these task features. Second, explicit, conscious, analytic processes are applied to these relevant task features in order for an inference or judgment to be made. Evans' (1998b) specific account of matching bias on the selection task is that it arises from the operation of a linguistically-based matching heuristic, which reflects the way in which negative terms are used in natural language to deny suppositions rather than to assert new information. The essential idea is that a negative statement is a comment that does not alter the topic of an assertion (e.g., a statement such as "there is not a 3" is still about the number 3 rather than any other number)¹. Evans also proposes that another heuristic, the if heuristic, arises in the context of conditional selection tasks, and causes attention to be focused on the true antecedent (TA) card for all rules. This heuristic explains why there is a good level of logical performance for TA cards (which should always be selected), and why the matching response is less marked for negated antecedent rules compared with affirmative antecedent ones.

Evans' account of selection task performance is, therefore, attentional in emphasis. Linguistic cues draw attention toward certain cards and away from others; the former get selected and the latter get rejected. Any analytic processing that is applied serves merely to rationalize decisions that have already been made on the basis of relevance (Evans, 1995). In this way, the H-A account can readily make sense of the finding

(Evans & Wason, 1976) that the retrospective verbal reports that people provide when asked to explain their card selections appear to reflect attempts to justify choices in terms of either verification or falsification (depending on the nature of the rule), with no apparent insight being displayed into the logical basis of selections.

Johnson-Laird and Byrne's (1991) mental models theory proposes that reasoning is based around the construction of models in which the premises of an argument or rule are represented as being true. To explain selection-task performance, this theory assumes that people: (1) only think about those cards that are explicitly represented in their models of the rule, and (2) only select those cards for which the hidden value could bear on the truth or falsity of the rule. So, for example, the failure to select the not-q card on If p then q reflects the fact that this term is not explicitly represented in the reasoner's models of the conditional, thereby resulting in the common selection of just the p card with this affirmative rule. Those people who represent this rule as a biconditional would select p and q (the other dominant selection combination for an affirmative rule), as both of these cards are explicitly represented in models and could have a bearing on the rule's truth or falsity. To account for matching bias on rules that contain negations, Johnson-Laird and Byrne (1991) argue that negated components promote the expansion of models to include the affirmative counterparts of negated terms. For example, If A then not 3 would be represented as:

[A] $\neg 3$

3

...

In this notation, each line depicts a different model which could be true if one assumes the truth of the conditional. The square brackets around 'A' in the first model denote that this proposition is exhausted with respect to $\neg 3$ (i.e., that whenever an A

occurs a $\neg 3$ also occurs). The second model is incomplete in that whether or not an A would occur given that there is a 3 has not been represented by the reasoner. In the third line, the ellipses denote an “implicit” model, meaning that there may be other models consistent with the rule to which the reasoner has not assigned explicit content. Central to the mental models theory of deduction is the notion that people have limited working-memory capacities, and hence will represent as little as possible in their initial model-set to capture the meaning of the connective if. Within the model-set for If A then not 3, it is only the A and the 3 cards whose hidden values could bear on its truth or falsity if it is treated as a conditional. It is, therefore, these cards that will be selected, so producing a matching response pattern.

The mental models theory of selection-task performance has some similarities to the H-A account (cf. Evans, 1998b). For example, the concept of explicit representation in models clearly overlaps with Evans’ (e.g., 1989) notion of relevance (i.e., what is explicitly represented in a model is what the individual perceives to be of relevance to the task at hand). However, the model theory differs critically from the H-A account in its assumption that a degree of analytic processing does determine card selections (i.e., the only cards that end up being chosen are those explicitly represented ones that are deemed to bear on the rule’s truth or falsity).

A third theory of selection-task performance, forwarded by Oaksford and Chater (e.g., 1994, 1996; see also Oaksford & Chater, 2003), is framed within their general rational-analysis approach to human reasoning, and is referred to as the optimal data selection (ODS) account. Oaksford and Chater propose that selections are based on the information value of cards in relation to their potential support for the rule, estimated in the form of expected information gain. Oaksford and Chater’s mathematical analysis of the information value of cards shows, for example, that the selection of the matching q

card for the affirmative conditional can be more useful than the selection of the non-matching (but logically appropriate) not-q card². In this way, the ODS model proposes that illogical matching choices may, in fact, be deemed to be rational in terms of a probabilistic standard.

The ODS theory presents a persuasive account of the matching effects observed on affirmative conditional rules within the selection task. Moreover, because the ODS theory capitalizes on Oaksford and Stenning's (1992) arguments that negations typically define high-probability contrast sets, it is also readily able to explain antecedent and consequent matching effects observed for conditional rules containing negated constituents (e.g., Oaksford, 2002a; Yama, 2001). So, for example, a rule such as If there is an A on one side of the card then there is not a 3 on the other side is argued to designate a high probability true consequent category (any number that is not a 3), whereas the false consequent category is represented by a very low probability single case (the matching 3 card), whose rarity assures its high information value.

Overall, then, the ODS account is able to accommodate a wide range of evidence for matching effects in the standard selection task paradigm. A final strength of the theory - and one which sets it apart from both the H-A and the mental models accounts -- is its capacity to explain the considerable body of evidence that has now been amassed for probabilistic influences on card selections (e.g., Green & Over, 1997; Green, Over, & Pyne, 1997; Kirby, 1994; Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, Grainger & Larkin, 1997). So, for example, it has been shown that card selections vary in ways predicted by ODS when P(p) and P(q) are varied experimentally. Non-probabilistic theories are generally not readily able to explain why probability manipulations should affect card selections, only really being able to do so by invoking the idea that participants adopt different task interpretations, with probabilistic

manipulations affecting the proportion of people adopting these different interpretations (see Oaksford & Wakefield, 2003).

In spite of the capacity of the ODS theory to explain an impressive range of selection-task data, it has been claimed to have certain limitations. One problem (cf. Evans, 2002) is the difficulty that the theory appears to have in explaining why the use of explicit negations on cards in selection tasks completely removes matching bias (e.g., Evans et al., 1996). This phenomenon is easily accounted for by the H-A theory, as all cards present matching values within an explicit negations paradigm. Oaksford (2002a), however, has recently proposed that this explicit negations effect may be a result of participants failing to engage their “normal” interpretative processes in this task variant -- an explanation that is certainly worthy of further investigation.

Tests of Theoretical Accounts of Matching Bias

Roberts (1998b) has commented that only limited attempts have been made to test directly theories of selection-task performance through the production of converging evidence beyond actual card-selection patterns. Moreover, what effort has been made in this respect has primarily been driven by researchers working within the H-A framework. For example, Evans, Ball and Brooks (1987) used computer-presented tasks and recorded the order in which “select-don’t select” decisions were made about each card. As predicted by the H-A account, people made decisions about matching cards before mismatching ones, and a correlation was found between card selection frequency and card decision order (i.e., selected cards were decided about earlier than rejected ones). However, it is possible that these results simply reflect a preference for people to register “select” choices before “don’t select” ones, and that what is being shown is a response bias as opposed to an attentional bias (Evans, Newstead & Byrne, 1993; Roberts, 1998b).

Evans (1996) provided stronger evidence for the H-A account in two studies using a mouse-pointing methodology to record card-inspection times. Participants were required to tackle computer-generated selection tasks, and to indicate cards they were “thinking about” by holding the mouse pointer over that card. Cards were selected by means of a mouse-click, whilst no action was required for non-selected cards. The computer logged cumulative inspection times for each of the four cards on a given problem. Evans argued that if heuristic processes were cueing card selections, then only such heuristically-cued cards would be subjected to analytic rationalization processes aimed at justifying their selection. Inspection-times would, therefore, be higher for selected cards than for rejected cards. The specific predictions that Evans tested were: (P1) cards associated with higher selection rates will be associated with longer inspection times, and (P2) for a given card, participants who choose it will have longer inspection times than those who do not. Evans (1996) found strong support for both predictions. There were often sizeable differences in mean inspection times between selected and non-selected cards, with the former generally being greater than 4 s and the latter typically being less than 2 s.

Despite this converging evidence for the H-A account of the selection task, Roberts (1998b) has noted a need for caution in interpreting findings that derive from Evans’ card-inspection paradigm. In particular, Roberts argued that there are three potential sources of bias inherent in this paradigm that could have led to artefactual support for the claim that people are spending time rationalizing choices that have been cued by relevance-determining heuristics. First, participants may have a tendency to pause the mouse pointer over a card for a brief period of time before making an active “select” decision about it (by clicking on it). This would lead to inflated inspection times for selected cards relative to non-selected ones. Second, participants may forget to move

the mouse pointer to a new card even though their attention has shifted to this new card. Since people have a known preference for registering “select” decisions before “don’t-select” decisions (Evans et al., 1987), such forgetfulness could, again, inflate the inspection times for selected cards (decided about earlier) than rejected cards (decided about later). Third, sensory leakage, where cards may be inspected (and even rejected) before the mouse pointer has had a chance to reach them, could also result in inflated times for selected over rejected cards.

Roberts (1998b) systematically manipulated the presence of these task-format biases across a series of experiments, and demonstrated that the magnitude of the inspection-time effect was closely related to the number of sources of bias present. With all sources removed the inspection-time effect was also eradicated. Moreover, in a unique deselection task -- where all cards were initially presented as selected, and participants clicked to deselect them -- a complete reversal of the inspection-time effect was observed (i.e., deselected cards were inspected for longer than those that remained selected). Overall, Roberts’ (1998b) results suggest that biases arising from task-format factors may provide at least as good an explanation of apparent inspection-time effects in the selection task as Evans’ H-A framework (see Evans, 1998a, and Roberts, 1998a, for a subsequent exchange of views on the implications of inspection-time findings for the H-A account).

More recently, Ball, Lucas, Miles and Gale (2003) have critiqued inspection-time studies using mouse-pointing because of their use of an inherently insensitive technique for monitoring the second-by-second transitions in attentional processing that arise during selection-task performance. Of particular concern is the fact that mouse pointing is an indirect measure of attentional processing since participants have actively and effortfully to move the mouse pointer to cards that they are thinking about. Ball et al.

advocate the use of eye-movement tracking as a more precise approach for measuring moment-by-moment attentional shifts that underlie cognitive performance with highly display-based problems like the selection task. Ball et al. report three experiments that systematically eradicated the sources of artefact discussed by Roberts (1998b) by combining careful task constructions with eye-movement tracking to measure directly on-line processing. All three experiments produced good evidence for the robustness of the inspection-time effect, so supporting the predictions of the H-A account.

A recent experiment by Roberts and Newton (2001, Experiment 1) that encompassed a methodological innovation to improve on previous mouse-pointing techniques also demonstrated a reliable association between card selection and increased inspection times, indicating that mouse pointing measures can be sensitive to effects predicted by the H-A framework. Roberts and Newton (2001), however, presented two further studies (Experiments 2 and 3) using a rapid-response selection task (requiring a card decision within 2 s of its presentation) that led them to propose an important caveat concerning the adequacy of the H-A theory, although they remain broadly favourable toward this account. They note that their rapid-response tasks raised levels of matching for consequent cards on certain rule forms (without increasing levels of logical responding) in comparison with free-time tasks. This suggests that analytic processing arising in free-time situations may serve to overturn candidate cards cued through attentional heuristics (in contradiction to Evans' H-A account, but in line with mental models proposals). It is possible, however, that the analytic effects that influence card selections identified by Roberts and Newton (2001) may be restricted to a subset of individuals, with a majority responding equivalently under both speeded and unspeeded conditions. Thus the H-A theory may capture the behaviour of most individuals, whilst other accounts (e.g., mental models theory) may better describe the processing of a

subset of individuals (see Stanovich & West, 1998, for evidence of individual differences in responding on the selection task).

Despite the general support for key tenets of the H-A account of selection task that derive from Roberts and Newton (2001) and Ball et al. (2003), both sets of researchers note that the magnitude of the inspection-time effect tends to be very small. Roberts and Newton observed that the difference between selected and non-selected cards was 0.3 s after transformation, a value that maps well onto the 0.36 s difference seen by Ball et al. in their Experiment 3 (which arguably involved the least influence of task-format biases). At first sight such data seem inconsistent with Evans' view that the inspection-time effect is attributable to conscious rationalization processes that are applied to to-be-selected cards with the aim of justifying heuristically-determined choices. Surely, analytic processes should take rather more than a fraction of a second to be applied in this way? Ball et al. (2003) note, however, that the idea that analytic rationalization should take a lengthy amount of time may be more apparent than real. In support of this assertion they point to research by Wason and Evans (1975) that investigated the justifications people provide for card selections. Wason and Evans uncovered a phenomenon termed secondary matching bias, which is a marked tendency for people to rationalize a card choice in terms of a matching value that might be present on the card's reverse side. Thus, for a rule such as If there is an A on one side of the card then there is a 3 on the other side, participants justify why the A card should be selected by stating how a 3 (a matching value) on the other side would verify the rule. It is entirely possible that any rationalization process, whilst analytic and evaluative in intent, may, nevertheless, be guided by the extremely rapid, heuristic cueing of relevant information.

Another possible account of the small inspection-time difference between selected and non-selected cards is that analytic rationalization processes do not actually occur at

all on the selection task (i.e., heuristic processing dominates all task-based activity). The inspection-time effect would, by this account, merely reflect the action of heuristic processes holding attention on relevant cards for slightly longer than irrelevant ones. As Ball et al. (2003) point out, however, under such an account it is unclear what kind of processing might be occurring during the extra period of heuristically-compelled attention to cards if analytic processing is believed not to be taking place.

Verbal-Protocol Data and the Selection Task

One approach to advancing a theoretical understanding of the processing arising during selection-task performance might be to obtain verbal “think-aloud” protocols from participants attempting such problems. The elicitation of verbal protocols from people engaging in problem solving and reasoning has become a respected method of cognitive enquiry since the publication of Ericsson and Simon’s (e.g., 1980, 1993) research assessing the validity of this approach. In particular, Ericsson and Simon (1993), having reviewed a wide body of pertinent literature, stress that the production of concurrent think-aloud reports can provide a highly accurate and complete index of the current contents of short-term memory, in that whatever is consciously attended to by a participant is also verbalisable. In addition, Ericsson and Simon (1993) provide compelling evidence indicating that retrospective reports elicited from people subsequent to task-based performance are problematic in terms of their validity, as they seem to be susceptible to biases arising from self-theorising on the part of the participant (cf. Nisbett & Wilson, 1977).

Given that concurrent think-aloud protocols appear to be effective for tracing the locus of attention during cognitive task performance in reasoning contexts (cf. Evans, 1989), it is perhaps surprising that there are few published reports that have employed this method in studying the selection task. Indeed some of the most recent research

using verbalisation methods have actually been at odds with Ericsson and Simon's recommendations for the importance of eliciting concurrent rather than retrospective reports. So, for example, Green and Larkin (1995) utilised a post-hoc reporting technique, whereby participants had explicitly to provide reasons for their card selections when prompted by the experimenter. Such focused, retrospective verbalisations may well tell us very little about the on-line focus of participants' moment-by-moment attentional processing as might be gleaned from the use of a concurrent verbalisation method. Another recent body of selection-task research by Stenning and van Lambalgen's (2002) elicited verbalisations from participants tackling selection tasks as part of a Socratic "tutorial dialogue" between experimenter and reasoner. As interesting as this methodology certainly is, we believe the technique may have only a limited bearing on the issue of individual reasoning processes divorced from the dynamics of didactic conversations between students and tutors. Indeed Stenning and van Lambalgen (2002, p. 281) themselves acknowledge that "Engaging subjects in dialogue undoubtedly changes their thoughts, and may even invoke learning. The relation between the reasoning processes evoked by the standard way of conducting the task, and the processes reflected in subsequent dialogues is a relation that remains to be clarified".

One key study of the selection task that has utilised more reliable concurrent-reporting methods is that reported by Beattie and Baron (1988, Experiments 2 and 3). This research provided evidence that participants rarely mentioned alternative cards to the ones that they ended up selecting -- an effect that Beattie and Baron viewed as supporting the notion of a heuristically-based matching process. Participants were also seen to be overconfident about their card choices and showed little sensitivity to the correctness of their selections. However, Beattie and Baron's protocol coding scheme

functioned at a fairly gross level of analysis that focused on the classification of selection patterns and the categorization of responses to probe questions. As such, their coding does not appear to have been geared toward uncovering insights into the spontaneous analytic processing that might be associated with card choices.

A more recent protocol analysis of the selection task was presented by Evans (1995, Experiment 5). Protocols were analysed in two distinct ways. First, they were scored for references to the facing sides of cards. The percentage references were then divided according to whether the participant selected the card or not. Second, protocols were scored for references to the hidden sides of cards, and again these scores were broken down according to whether the card was selected or not. Consistent with Beattie and Baron's (1988) findings, Evans' first analysis revealed a significant and substantial tendency for participants to refer more often to the facing sides of cards that ended up being selected than to the facing sides of cards that ended up being rejected. Importantly, however, Evans' secondary analysis revealed an identical tendency for participants to refer more to the hidden sides of selected cards than to the hidden sides of non-selected cards. Evans (1995) viewed these findings as supporting the H-A position that people think only about some cards and not others, and that thinking about hidden sides of cards mostly serves to rationalize decisions to choose such cards.

Experiment 1

We are generally persuaded by Evans' (1995) protocol-based support for the role of relevance effects and rationalization processes in the selection task. It is noteworthy, however, that Evans' findings derive from the analysis of four distinct experimental conditions involving very small sample sizes (i.e., ns of 3, 3, 4, and 5). In addition, Evans' tasks involved arbitrarily thematic selection-task materials (as opposed to abstract contents), and certain experimental conditions entailed highly non-standard

judgement instructions. Furthermore, Evans' statistical analysis of his dataset using chi-squared tests was potentially problematic in that participants contributed multiple data points to both the selected and non-selected cells of the contingency tables. Finally, Evans' analyses did not focus on the important issue of the content of people's references to potential values that may reside on the hidden sides of cards. Gaining an understanding of whether secondary matching bias effects (Wason & Evans, 1975) are associated with hidden-side references would be especially valuable for advancing an understanding of why the attentional processing of selected cards evidenced in inspection-time studies (Ball et al., 2003; Roberts & Newton, 2001) seems to be increased only to a small (though reliable) degree relative to non-selected cards.

Overall, then, there would seem to be clear scope for replicating Evans' (1995) protocol-based findings with an increased sample size, more conventional task features -- including the employment of standard abstract problems -- and traditional task instructions. Such a replication formed the primary aim of our first experiment. Crucially, too, we set out to adapt the H-A predictions that have been applied effectively in inspection-time studies (e.g., Ball et al., 2003; Evans, 1996; Roberts, 1998b) so as to enable more powerful statistical tests to be pursued of the H-A theory in terms of people's references to the facing and hidden sides of selected and non-selected cards. To this latter end we established three key predictions:

P1 -- Cards that are associated with higher selection rates will also be associated with more references to their facing sides.

P2 -- For any given card, those participants who select it will refer more to its facing side than those participants who do not select it.

P3 -- For each participant, their mean number of references to the facing sides of selected cards should be higher than to the facing sides of non-selected cards.

The latter participant-level prediction is a variant of that advocated by Roberts (1998b) in the context of card inspection-time analyses, and is argued to be a more powerful test of the H-A account than either P1 or P2, which involve item-level analyses. It should be noted, too, that all three predictions have been stated solely in terms of references to the facing sides of cards. It is also possible, however, to restate each of these predictions so that they apply equally to the analysis of references to the hidden sides of cards. Such re-stated predictions would be entirely in line with the claim of the H-A theory (e.g., Evans 1995, 1996) that rationalization processes serve merely to justify card choices, thereby promoting increased references to hidden sides of selected cards relative to the hidden sides of non-selected ones.

Finally we derived one further prediction from the H-A theory that pertained to the content of people's explicit references to the hidden sides of cards. This prediction was as follows:

P4 -- The total pool of references to hidden sides of cards should be dominated by references to potential matching values that might appear on the reverse sides of cards relative to either mismatching values or negated matching values.

This last prediction derives from the assumption that secondary matching heuristics may guide the analytic rationalization processes associated with to-be-selected cards (Wason & Evans, 1975; Ball et al., 2003).

Method

Participants. Participants were 30 undergraduate volunteers from the University of Derby who took part in the experiment to gain course credit. Participants had not received any tuition on the psychology of reasoning.

Materials and apparatus. The experiment involved selection tasks employing abstract conditional rules within a standard negations paradigm. Each participant

received four versions of the task, with negatives permuted through the conditional statements as shown in Table 1a (conventional terminology is used throughout the paper to refer to cards). Each problem was presented on a single A4 page. The rule was positioned at the top of the page, a reminder of the task requirement appeared in the middle of the page, and the pictures of the four cards were presented in the lower half of the page in a two-by-two arrangement. The location of cards within each array was always random. The experiment was carried out in an audio-recording suite to enable participants' think-aloud protocols to be recorded.

(Table 1 about here)

Procedure. Participants were tested individually. They were initially told about the essential nature of the experiment and the basic “think-aloud” requirement. To help clarify the expectations surrounding the think-aloud requirement and to put participants at their ease a brief video-based demonstration was provided of someone verbalizing whilst carrying out a moderately difficult problem-solving task involving the rebuilding of a pyramid structure using jigsaw-like building blocks. Subsequent to this demonstration the following written instructions were presented:

This study is concerned with people's logical reasoning ability and will entail you having to tackle a total of four problems. These problems will appear on separate sheets in front of you. Each problem consists of four cards and a rule that applies to those cards. This rule may be true or false. The cards have been constructed so that each one always has a letter on one side and a single-figure number on the other side. Naturally only one side of each card will be visible to you.

For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. It is all right for you to

change your mind as you work through a problem, and I will not record any decisions until you tell me what your final choice or choices are.

Whilst you are reading through each problem and deciding how to solve it, please remember that I would like you to think aloud. As I've explained, you should find it quite natural to say aloud whatever happens to come into your head whilst you are working on these tasks. If you do fall silent for any length of time, however, I will gently prompt you to try and keep thinking aloud.

Once the participant had read the instructions the experimenter re-read them aloud and provided an opportunity for participants to seek clarification concerning any of the study requirements. The four problems were then presented in a random order.

Results

Card-selection patterns. Our first concern was to assess whether our tasks elicited the standard pattern of card selections observed in the literature (i.e., more matching than mismatching choices across antecedent and consequent cases). Matching bias was examined using the procedures adopted by Evans et al. (1987), summarized in Table 1b. An alpha level of .05 was set for all tests reported throughout this paper. Wilcoxon signed-rank tests (one-tailed) revealed that more antecedent matching cards were selected than antecedent mismatching ones, $p = .01$, and that more consequent matching cards were selected than consequent mismatching ones, $p < .001$. This pattern of results is, therefore, typical of that seen for selection tasks within the negations paradigm.

Protocol coding, reliability assessment, and normality checks. Verbal protocols were transcribed and then coded using three categorization systems. The first scheme was inspired by Evans (1995, Experiment 5) and involved examining an individual's protocol and, for each rule, identifying the discrete references to the facing side of each card. Frequency counts for each participant's total number of references per card were

then calculated in order to provide a measure for use in subsequent statistical analyses. It is important to note that in applying this first scheme we were cautious not to code any references to facing card sides that occurred when participants were making or confirming their final card selections. This meant that we could avoid the possibility of obtaining artefactual support for H-A predictions arising from the fact that only selected cards needed to be actively registered by participants. Any references to facing sides arising at the selection-registering phase would artificially inflate the frequency-count of mentions to selected cards, since it is only these cards that need to be referred to explicitly (for similar methodological concerns relating to the inspection-time effect in selection task studies see Roberts, 1998b, and Ball et al., 2003). In essence, our conservative measure of references to facing sides provides a stronger test of H-A predictions than the coding scheme applied by Evans (1995, Experiment 5), which appears not to have considered such methodological artefacts.

Our second categorization scheme was identical to the previous one in all respects, except for its focus on a participant's references to the hidden side of each card (again, see Evans, 1995, Experiment 5). Two coders (the authors of the present paper) independently applied both of the aforementioned categorization schemes to the full set of verbal protocols. Inter-coder reliability checks revealed a very high degree of consistency between coders (i.e., 97% inter-coder agreement), and there was no evidence of systematic divergences between coders in their categorization of discrete references to the facing or hidden sides of each logical case. The codes applied by the second author were used for all analyses associated with P1, P2 and P3.

Our third categorization scheme involved sub-categorising each reference to a hidden side in terms of the specific letter or number content mentioned in that reference. This coding scheme used the following four sub-categories, which are illustrated in

terms of participants' potential references to what might have been on the other side of the A card associated with an If A then 3 rule: (1) a reference to a matching item (e.g., mentioning the possibility of a 3 on the other side of the A card); (2) a reference to mismatching item (e.g., mentioning the possibility of a number such as a 7, on the other side of the A); (3) a reference to a negated matching item (e.g., stating that there could be a number that is not a 3 on the reverse of the A); and (4) a non-specific reference to what might be on the other side of the card (e.g., when participants stated how "It is important to see what's on the other side of the A", without qualifying such comments further). It should be noted that whilst other sub-categories are possible in addition to the four we have described (e.g., references to negated mismatching items), our four sub-categories successfully captured the full range of content that we discerned in participants' references to the hidden sides of cards. This third coding scheme was applied by the second author. As there was only limited scope for miscategorizing references using this scheme (i.e., the new codes simply reflected a more detailed breakdown of the explicit references to hidden card sides that had already been identified) we did not deem it necessary to pursue inter-coder reliability checks on the application of these codes.

Descriptive analysis of our data revealed that they were positively skewed. Log transformations (subsequent to the addition of suitable constants) were found to stabilize variances successfully. Identical problems with skew were encountered in Experiment 2, and statistical analyses were, therefore, performed on log-transformed data throughout both experiments³. For clarity of interpretation we report means both before transformation and converted back into their original units after transformation.

Analyses relating to H-A predictions. Our statistical analyses examined the four predictions, identified above, that derive from Evans' (e.g., 1996) H-A account. As

noted previously, P1, P2 and P3 apply equivalently to the measure of references to the facing sides of cards and to the measure of references to the hidden sides of cards.

Our first analysis tested P1: Cards associated with higher selection rates will be associated with more references to their facing (and hidden) sides. This analysis involved exploring the correlation across all 16 cards between the overall mean references to a card side and the card's associated selection frequency (refer to Table 2). The correlation between the mean number of references to facing sides and card selection frequencies revealed a strong positive association, $r = .94$, $N = 16$, $p < .001$ (transformed data). The correlation between mean references to hidden sides and selection frequencies was also positive and highly reliable, $r = .89$, $N = 16$, $p < .001$ (transformed data).

(Table 2 about here)

The second analysis tested P2: For any given card, mean references to a card side for individuals selecting it would be higher than for those who did not select it. Mean references to facing and to hidden sides for each card are given in Table 3. After transformation, mean references to facing sides for 16 out of 16 cards were greater for selectors than non-selectors, $p < .001$, two-tailed binomial test. Likewise, after transformation, mean references to hidden sides for 15 out of 16 cards (one tie) were greater for selectors than non-selectors, $p = .001$, two-tailed binomial test⁴.

(Table 3 about here)

P3 tested whether, for each individual, the mean references to card sides were higher for selected than for non-selected cards. Two mean references to facing sides scores and two mean references to hidden sides scores were calculated for each person from the transformed data (Table 4). A within-participants analysis of variance

(ANOVA) provided strong support for P3 for facing sides, $F(1,29) = 115.44$, $MSE = 2.54$, $p < .001$, as well as for hidden sides, $F(1,29) = 106.43$, $MSE = 2.13$, $p < .001$.

(Table 4 about here)

We assessed secondary matching bias predictions pertaining to P4 by taking the total pool of references to hidden sides of cards produced by all 30 participants, and then computing the distribution of references within this pool across the four sub-categories of reference-type (i.e., matching items, mismatching items, negated matching items, and non-specific references). This analysis revealed that the mention of matching values dominated people's verbalizations concerning what might appear on the reverse sides of cards (64% of references) in relation to the mention of negated matching values (35% of references), mismatching values (< 1% of references) and unspecified values (< 1% of references).

Discussion

The results of Experiment 1 support predictions that can be derived from Evans' (e.g., 1996) H-A account of the selection task. Overall, the analyses demonstrate that selected cards are associated with more attention than non-selected cards, as indexed by the quantity of explicit references to both their facing sides and their hidden sides. The finding that H-A predictions are substantiated in the analysis of references to the hidden sides appears to lend support to Evans' claimed role for analytic rationalization processes during selection-task performance. Thus, whilst to-be-selected cards elicit increased consideration of what values might be on their reverse sides, such consideration seems to do little to change the fact that these cards tend to end up being selected (otherwise the link between references to hidden sides and card selection would be broken). One way to interpret our findings, then, is that thinking about the hidden sides of cards appears to have a minimal functional role in determining card

choices (at least for a substantial number of participants), instead mainly serving to confirm decisions to go ahead and choose such cards (cf. Evans, 1995, p. 168).

The issue of what people are actually thinking about when they consider the reverse sides of cards has also been addressed in the present study, with some clear-cut findings. First, participants do not think about potential mismatching values that may appear on the reverse sides of cards. This may be taken as further support for the H-A view that people tend not to see mismatching values as having any relevance to their decision making during the evaluation of conditional statements. Second, the finding that people's consideration of hidden values is dominated by matching possibilities seems to be in line with Wason and Evans' (1975) notion that secondary matching heuristics may cue people's analytic accounts as to why values on the hidden sides of cards justify selection of those cards. This evidence for secondary matching effects in abstract selection tasks also helps make sense of card inspection-time findings (e.g., Ball et al., 2003) that suggest analytic rationalization processes are rapid in nature. Rationalization might well be expected to be extremely fast if people's justifications are facilitated by the heuristic cueing of "relevant" (i.e., matching) values that could appear on the reverse sides of cards.

Although we believe the findings from Experiment 1 are congruent with Evans' H-A account of the selection task, it is possible that other contemporary theories such as the ODS and mental models accounts could also accommodate these results. We return to this pertinent issue after first presenting our Experiment 2 data.

Experiment 2

Experiment 1 produced findings that are both in line with predictions of the H-A theory of the selection task, and that also converge with evidence previously obtained from mouse-tracking and eye-tracking studies of card inspection times (e.g., Evans,

1996; Ball et al., 2003) and from small-scale protocol studies reported by Evans (1995). In addition, Experiment 1 has clarified the important role played by secondary matching processes when people are referring to the hidden sides of to-be-selected cards. One issue that remains open to further examination, however, is the effect that enforced attention to cards might have on people's heuristic and analytic processing, as measured by both the frequency of references to facing and hidden sides of selected and non-selected cards, and the content of references to hidden sides. A number of researchers have looked at the effects of enforced attention to all four cards in abstract selection tasks, including Ball et al. (2003, Experiment 3), Evans et al. (1987), and Roberts (1998b, Experiment 3). These studies required participants to provide "select-don't select" responses for all cards, and, in every case, it was found that matching-based responses continued to dominate selections despite the fact that normally unattended cards had to be responded to. Importantly, too, Ball et al. (2003) demonstrated that the inspection-time effect (whereby selected cards are looked at for longer than rejected cards) was not totally undermined by this "select-don't select" decision requirement, although the effect was reduced in magnitude, presumably because rejected cards now become associated with at least some (enforced) consideration.

A critical question that arises from the enforced-decision paradigm, however, is what, exactly, do people think about when they are compelled to inspect cards that they would not ordinarily attend to? In particular, do people who are making card selections within this paradigm think beyond the facing sides of those cards that they choose to reject? The H-A theory would surely still argue that people should not think about what might appear on the hidden sides of to-be-rejected cards. Although the enforced decision requirement means that people must attend to such cards in order to register active "don't select" responses to them, the fact that these cards should rapidly be

deemed to be irrelevant means that analytic rationalization processes would not be called upon (cf. Evans 1998a). In other words, rationalization processes in the selection task (and perhaps more generally too) are assumed to be “asymmetrical”, in that people only pursue analytic justifications for cards that they wish to select (as cued by relevance), but not for cards which they wish to reject (on the basis of perceived irrelevance). Such avoidance of any further processing of information deemed as “irrelevant” certainly makes sense on the grounds of efficient information processing (see Evans, 1983, for further discussion of the selective nature of reasoning).

In spite of the psychological plausibility of the arguments concerning selective information processing that derive from the H-A theory, it remains the case that the theory’s specific predictions pertaining to the enforced decision paradigm need to be assessed empirically. Experiment 2 was undertaken to test such predictions using the same selection tasks and think-aloud instructions employed in Experiment 1, except for the presence of enforced “select-don’t select” requirements for all cards. To test the H-A theory we employed the equivalent predictions for both facing and hidden card sides as we had used in Experiment 1. Although we expected some possible weakening of effect sizes for the facing side predictions (P1, P2, and P3) owing to the enforced decision procedure, our previous inspection-time data (Ball et al., 2003, Experiment 3) gave us grounds for assuming that the basic finding of increased attention to selected cards over rejected ones should remain intact (i.e., people would give to-be-rejected cards only a minimal amount of explicit consideration, dwelling instead on to-be-selected cards). As far as our predictions for the hidden sides of cards were concerned, we predicted effects of broadly similar magnitude to those that arose in P1 to P4 of Experiment 1 (i.e., participants were not expected to think about the reverse sides of to-be-rejected cards any more than in the standard selection-task paradigm).

Method

Participants. Participants were 30 undergraduate volunteers from the University of Derby, who obtained course credit for their involvement in the study. No participants had received prior tuition concerning the psychology of reasoning.

Materials and apparatus. The selection-task materials and apparatus were identical to those used in Experiment 1, with the minor addition of small “yes” and “no” decision boxes that appeared approximately 1 cm below each card.

Procedure. The procedure was identical to Experiment 1, with the exception that instructions were modified to include reference to the presence of “yes” and “no” response boxes below cards. The second paragraph of the instructions was, therefore, amended to read as follows:

For each problem your task is to decide which card or cards need to be turned over in order to discover whether or not the rule is true. You will need to make a “turn-don’t turn” decision about all the cards presented to you. The “yes” and “no” boxes underneath each card are present to remind you that you must make a “turn-don’t turn” decision for every card. It is all right for you to change your mind as you work through a problem, and I will not record any decisions until you tell me what your final answers are for each card.

As in Experiment 1, once the participant had read the instructions the experimenter then read them aloud once more to enable any clarification to be sought concerning the task requirements. The four problems were presented in a random order.

Results

Card-selection patterns. Wilcoxon signed-ranks tests (one-tailed) revealed that more antecedent matching cards were selected than antecedent mismatching ones, $p < .001$, and that more consequent matching cards were selected than consequent mismatching

ones, $p < .001$. The standard matching-bias pattern is, therefore, strongly evident in the card-selection responses associated with this enforced decision paradigm.

Protocol coding, reliability assessment, and normality checks. Transcribed protocols were coded using identical categorization and scoring schemes as had been applied in Experiment 1. Inter-coder reliability checks revealed a high level of consistency between coders in their application of the categorization schemes pertaining to references to facing and to hidden sides of cards (i.e., 95% inter-coder agreement). The codes applied by the second author were used for all subsequent analyses associated with the experimental predictions (P1 to P4). Similar data transformations to those used in Experiment 1 were employed to overcome positive skew and stabilize variances.

Analyses relating to H-A predictions. The mean number of references to the facing and the hidden sides of each card, and each card's overall selection frequency, are presented in Table 5. The correlations for P1 between mean references to card sides and selection frequencies were significant for facing sides, $r = .88$, $N = 16$, $p < .001$ (transformed data), and for hidden sides, $r = .94$, $N = 16$, $p < .001$ (transformed data).

(Table 5 about here)

In relation to P2, mean references to the facing and to the hidden sides of each card for selectors and for non-selectors are given in Table 6. For transformed data, the difference between mean references to facing sides for selectors and non-selectors was in the expected direction for 13 of the 16 cards (two ties), which was significant with a binomial test, $p = .021$, two-tailed. The difference between mean references to hidden sides for selectors and non-selectors was in the expected direction for 15 out of 16 cases, which was highly reliable with a binomial test, $p = .001$, two-tailed.

(Table 6 about here)

To assess P3 we again undertook more powerful participant-level analyses using ANOVA, which revealed (see Table 7) a significant difference in the mean references to facing sides for participants' selected versus non-selected cards, $F(1,29) = 4.62$, $MSE = .04$, $p = .04$, and a significant difference in the mean references to hidden sides for participants' selected versus non-selected cards, $F(1,29) = 8.94$, $MSE = .40$, $p = .006$.

(Table 7 about here)

Finally, we assessed secondary matching bias predictions associated with P4 by calculating the distribution of all participants' references to hidden sides across the four sub-categories of reference-type: matching items, mismatching items, negated matching items, and non-specific references. The mention of matching values dominated participants' comments about what might appear on the reverse sides of cards (62% of references) in relation to the mention of negated matching values (33% of references), mismatching values and unspecified values (< 3% of references in each case). This distribution of references to hidden sides across these four categories is strikingly similar to the distribution observed in Experiment 1.

Discussion

The results of Experiment 2 were, again, in line with all the predictions of the H-A account of performance on abstract versions of the selection task. As in Experiment 1, people referred more to the facing and hidden sides of those cards that ended up being selected relative to those cards that ended up being rejected. This evidence for the H-A theory arises in spite of the fact that the enforced decision paradigm necessarily requires people to give at least some attention to cards that they might ordinarily simply ignore on the basis of their perceived irrelevance to the task at hand.

Prior to undertaking Experiment 2 we had also speculated about possible shifts in the effect sizes for the facing and hidden side predictions (P1 to P3). More specifically,

we anticipated that the employment of enforced decision instructions might have some weakening influence on the magnitude of the effect size for the facing side predictions, since participants had to give some attention to all cards. On the other hand we had assumed that there should be no real changes in the magnitude of the effect size for the hidden side predictions, since participants were not expected to go on to think any more about the reverse sides of to-be-rejected cards even if forced to consider their facing sides. As it turned out, we were essentially correct in our expectations about changes to the effects we had observed in Experiment 1. So for example, in relation to the P2 item-based analysis, whereas 16 out of 16 cards in Experiment 1 showed increased references to facing sides for selectors compared to non-selectors, this dropped slightly to 13 out of 16 cards in Experiment 2. In contrast, there was no such drop between Experiments 1 and 2 in terms of references to hidden sides for selectors compared to non-selectors across cards (i.e., 15 out of 15 cards showed expected differences in both experiments).

A broadly similar pattern of changes to effect magnitudes was seen across Experiments 1 and 2 in relation to the P3 participant-based analyses. From Tables 4 and 7 it can be seen that the mean difference in references to facing sides for selected versus non-selected cards dropped quite markedly from 1.02 references in Experiment 1 (i.e. 1.26 minus 0.24) to 0.20 references in Experiment 2 (i.e., 1.08 minus 0.88), whereas the mean difference in references to hidden sides for selected versus non-selected cards dropped less strikingly from 0.66 references (i.e., 0.73 minus 0.07) in Experiment 1 to 0.17 references (i.e., 0.34 minus 0.17) in Experiment 2.

General Discussion

Experiments 1 and 2 were motivated by Evans' (e.g., 1996) H-A account of matching-bias effects with abstract selection tasks. This account claims that

preconscious, heuristic processes direct attention towards cards that appear to be relevant (which end up being selected) and away from ones that appear to be irrelevant (which are then rejected). Moreover, any conscious analytic processing that is applied to cards is assumed to have little functional role in determining selections, serving instead to rationalize decisions that have already been made on the basis of relevance judgements. In our experiments we elicited concurrent verbal protocols from individuals tackling selection tasks in order to examine H-A predictions concerning what people think about when deliberating over cards. The think-aloud method is a valuable way to study sequential thinking activity as it is typically viewed as having the capacity to provide a reliable index of the locus of participants' attentional focus during task performance (Ericsson & Simon, 1993; Evans, 1989). Although the think-aloud method may be less sensitive than other methods (e.g., eye-movement tracking) for monitoring the second-by-second shifts that arise in the focus of attention, a key advantage of the approach is that it provides an explicit trace of the content of people's thoughts about those task components that have captured attention.

Experiment 1 used a standard selection-task paradigm where active, select decisions were required only for cards that participants thought should be turned over to test the given rules. The protocol-based evidence obtained in Experiment 1 was directly in line with predictions derivable from the H-A theory. Participants referred reliably more often to the facing and the hidden sides of those cards that they selected compared with those that they rejected. These results substantiate Evans' (1995) findings from a set of small-scale protocol studies involving low participant numbers and a variety of non-standard instructional and contextual factors. In addition, Experiment 1 revealed new evidence for the role of secondary matching biases (Wason & Evans, 1975) dominating people's references to the hidden sides of cards. This latter finding suggests that

people's analytic processes may be supported by the rapid, secondary cueing of matching information. This result also clarifies why the card inspection-time data of Ball et al. (2000; see also Roberts & Newton, 2001) indicate that even selected cards are associated with quite minimal processing effort. That is, if analytic rationalization processes are supported by the rapid, heuristic cueing of "relevant" values that might arise on the reverse sides of cards, then there is no reason to expect such rationalizations to take more than the briefest amount of time.

Experiment 2 employed an enforced choice selection-task paradigm to assess what impact placing a "select-don't select" decision requirement on all cards might have on the content of people's thinking. We know from previous research (e.g., Ball et al., 2003, Experiment 3) that matching-bias and card-inspection-time effects remain intact despite such enforced decision requirements. However, the fact that Ball et al. observed a reduction in the magnitude of the predicted inspection-time effect suggests that compelling people to attend to all cards might also have a small but detectable impact on the effect magnitudes for H-A predictions pertaining to references to facing sides of cards. This was indeed seen to be the case. All facing side predictions were supported, but there was some evidence of reduction in the size of the observed effects. Perhaps more importantly, however, in the case of H-A predictions pertaining to references to hidden sides of cards, we expected reliable effects (as in Experiment 1), but with no particularly marked impact on effect magnitudes. This was because the H-A theory would argue that people should not think about what values might be on the hidden sides of to-be-rejected cards (since these are judged to be irrelevant), even if the task instructions necessitate that people have to attend momentarily to the facing sides of such cards. Again, all H-A expectations gained support from the protocol-based data obtained in Experiment 2, with reliable analytic-processing effects in evidence for

selected cards versus rejected cards, and less noticeable reductions in effect magnitudes for the hidden side predictions compared with the facing side predictions.

Overall, then, we believe that we have uncovered protocol-based evidence for the role of relevance effects influencing both heuristic and analytic processing in abstract selection task performance, as predicted by Evans' H-A theory. It is important, however, to consider whether other contemporary theories of the selection task are able to accommodate the present set of findings. Indeed, we acknowledge that our study did not encompass a crucial experiment that was set up to arbitrate unequivocally between different theoretical perspectives on the selection task. Thus it may well be that whilst our findings are congruent with the H-A theory that motivated our research, they may be similarly amenable to interpretation by one or more other contemporary selection task theories.

As we noted previously, Oaksford and Chater's (e.g., 1994, 1996) ODS account has a compelling track record in terms of its capacity to explain many aspects of selection-task performance (including the influence of probabilistic manipulations) across a variety of task variants. According to the ODS theory, information gain provides a formal measure of "relevance" (see Oaksford & Chater, 1995), and therefore ODS predicts the same basic pattern of matching-card selections as envisaged by the H-A theory. Moreover, ODS theory seems readily able to provide an alternative H-A account of our protocol-based findings that is distinct from Evans' emphasis on the linguistic basis of matching effects (we are grateful to Oaksford, personal communication, for alerting us to this). So, for example, relevance assessments determined on-line by participants via information-gain calculations could lead to more references to matching versus mismatching values on facing sides. More crucially, however, ODS theory would also predict that in justifying their card selections participants would show

secondary matching because for all rules they are searching for the rare case (which is always the matching antecedent and matching consequent combination).

Overall, then, the ODS model is able to capture the relevance effects that we have demonstrated in relation to references to both facing and hidden card sides, and, moreover, describes secondary matching as an analytic response (Oaksford, personal communication). This account also ties in selection task behaviour to rational explanations of biases in judgements relating to 2 by 2 contingency tables (e.g., Anderson & Sheu, 1995; Over & Green, 2001). One potential weakness with the ODS account as it is currently formulated is that it does not provide a full-blown algorithmic level theory specifying the specific nature, organization and time-course of the processing steps underpinning card selections (i.e., it is formulated at the computational level of what is being computed). As Oaksford (personal communication) has pointed out, however, most current models of the selection task (and not just the ODS theory) are actually highly underspecified in terms of the detailed operation sequences underlying reasoning, with theories tending merely to invoke a binary processing distinction (e.g., heuristic then analytic; initial representation then fleshing out). Indeed, the ODS model seems to be at least as capable as other theories of affording an understanding of algorithmic level issues in the abstract selection task -- as has been outlined above. Nonetheless, it would be appealing to see the ODS theory developed further at an algorithmic level; we understand that such developments are currently underway (e.g., Oaksford, 2002b), and look forward to their fruition.

The mental models account of the selection task (e.g., Johnson-Laird, 1995) may also be able to lend itself to an explanation of some of the key findings arising in the present experiments. One way to formalize an analysis of mental models predictions concerning people's explicit references to card sides is to compare the mean number of

references that people make to the sides of those cards that should be explicitly represented in models with the mean number of references to the sides of those cards that should not be explicitly represent in models. The key prediction (P5) would be that cards represented in mental models should be associated with more references to their facing and their hidden sides than cards not represented in models (see Ball et al., 2003, for equivalent analyses in relation to card-inspection time data).

We tested P5 for Experiment 1 by deriving participant-based scores for mean references to facing sides and mean references to hidden sides (see Table 8a). Repeated measures ANOVAs revealed good support for P5 for references to facing sides, $F(1,29) = 44.20$, $MSE = 0.56$, $p < .001$, and references to hidden sides: $F(1,29) = 35.72$, $MSE = 0.53$, $p < .001$, although it is noticeable that the effect magnitudes for P5 are less than those deriving from the equivalent by-participants analysis (P3) that investigated the H-A theory. We also conducted the P5 analysis for our Experiment 2 dataset (see Table 8b). Again, repeated measures ANOVAs revealed good support for P5 for references to facing sides of cards, $F(1,29) = 15.88$, $MSE = 0.21$, $p < .001$ and for hidden sides of cards, $F(1,29) = 35.03$, $MSE = 0.90$, $p < .001$. Unlike Experiment 1, the effect magnitudes associated with the test of the mental models predictions (P5) now appeared to be larger than those for the H-A predictions (P3).

(Table 8 about here)

Although it appears that the mental models theory can provide a coherent account of much of our protocol-based data, there remains a crucial set of evidence deriving from Experiments 1 and 2 that appears to arbitrate in favour of a H-A or ODS interpretation of performance on the selection task. This evidence concerns the finding that people's references to the hidden sides of cards are dominated by the mention of matching values over other possible entities. This apparent asymmetry in what values people consider as

being present on the reverse sides of cards does not readily seem to emerge from the mental models assumption that people assess cards in terms of how their hidden values might impact on the truth or falsity of the presented rule. It may well be that mental models theorists could develop a viable account for such secondary matching effects, but it remains the case that these effects were directly predicted by the H-A theory.

Our secondary matching evidence also challenge Feeney and Handley's (e.g., 2000) claims to have detected a deductive component in abstract variants of the selection task -- a conclusion they base solely on their finding that participants consider the hidden sides of presented cards. However, if when considering such hidden values most people are simply engaging in a secondary matching process, then this would seem to be evidence against deduction being a key component of reasoning in the selection task⁵. The limited support for mental models predictions deriving from evidence for secondary-matching effects also calls into question Evans' past proposals (e.g., Evans & Over, 1996, p. 136) that mental modelling may supply the analytic component to the H-A theory, which has always been less well specified than the heuristic component in this account. On balance, it would seem that either Evans' H-A account (minus a mental-models analytic stage) or Oaksford and Chater's ODS theory are most readily able to explain the full breadth of protocol-based evidence that we have uncovered for relevance effects and rationalization processes in the selection task. ODS theory may well have the edge on the H-A account, however, because of its impressive ability to explain a wide range of probabilistic influences on card-selection patterns.

References

- Anderson, J.R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. Memory and Cognition, 23, 510-524.
- Ball, L.J., Lucas, E.J., Miles, J.N.V., & Gale, A.G. (2003). Inspection times and the selection task: What do eye-movements reveal about relevance effects? Quarterly Journal of Experimental Psychology, 56A, 1053-1077.
- Beattie, J., & Baron, J. (1988). Confirmation and matching biases in hypothesis testing. Quarterly Journal of Experimental Psychology, 40A, 269-297.
- Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. Psychological Review, 87, 215-251.
- Ericsson, K.A., & Simon, H.A. (1993). Protocol analysis: Verbal reports as data (2nd edn.). Cambridge, MA: MIT Press.
- Evans, J.St.B.T. (1983). Selective processes in reasoning. In J.St.B.T. Evans (Ed.), Thinking and reasoning: Psychological approaches (pp. 135-163). London: Routledge and Kegan Paul.
- Evans, J.St.B.T. (1989). Bias in human reasoning: Causes and consequences. Hove: Lawrence Erlbaum Associates
- Evans, J.St.B.T. (1995). Relevance and reasoning. In S.E. Newstead and J.St.B.T. Evans (Eds.), Perspectives on thinking and reasoning: Essays in honour of Peter Wason (pp. 147-172). Hove: Lawrence Erlbaum Associates.
- Evans, J.St.B.T. (1996). Deciding before you think: Relevance and reasoning in the selection task. British Journal of Psychology, 87, 223-240.
- Evans, J.St.B.T. (1998a). Inspection times, relevance and reasoning: A reply to Roberts. Quarterly Journal of Experimental Psychology, 51A, 811-814.

Evans, J.St.B.T. (1998b). Matching bias in conditional reasoning: Do we understand it after 25 years? Thinking and Reasoning, 4, 45-82.

Evans, J.St.B.T. (2002). Matching bias and set sizes: A discussion of Yama (2001). Thinking and Reasoning, 8, 153-163.

Evans, J.St.B.T., Ball, L.J., & Brooks, P.G. (1987). Attentional bias and decision order in a reasoning task. British Journal of Psychology, 78, 385-394.

Evans, J.St.B.T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. Journal of Memory and Language, 35, 392-409.

Evans, J.St.B.T., Legrenzi, P., & Girotto, V. (1999). The influence of linguistic form on reasoning: The case of matching bias. Quarterly Journal of Experimental Psychology, 52A, 185-216.

Evans, J.St.B.T., & Lynch, J.S. (1973). Matching bias in the selection task. British Journal of Psychology, 64, 391-397.

Evans, J.St.B.T., & Newstead, S.E. (1980). A study of disjunctive reasoning. Psychological Research, 41, 373-388.

Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). Human reasoning: The psychology of deduction. Hove: Lawrence Erlbaum Associates.

Evans, J.St.B.T., & Over, D.E. (1996). Rationality and reasoning. Hove: Psychology Press.

Evans, J.St.B.T., & Wason, P.W. (1976). Rationalization in a reasoning task. British Journal of Psychology, 67, 479-486.

Feeney, A., & Handley, S.J. (2000). The suppression of 'q' card selections: Evidence for deductive inference in Wason's selection task. Quarterly Journal of Experimental Psychology 53A, 1224-1242.

Green, D. W., & Larkin, R. (1995). The locus of facilitation in the abstract selection task. Thinking and Reasoning, 1, 183-199.

Green, D.W., & Over, D.E. (1997). Causal inference, contingency tables and the selection task. Current Psychology of Cognition, 16, 459-487.

Green, D.W., Over, D., & Pyne, R. (1997). Probability and choice in the selection task. Thinking and Reasoning, 3, 209-236.

Johnson-Laird, P.N. (1995). Inference and mental models. In S.E. Newstead and J.St.B.T. Evans (Eds.), Perspectives on thinking and reasoning: Essays in honour of Peter Wason (pp. 115-146). Hove: Lawrence Erlbaum Associates.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). Deduction. Hove: Lawrence Erlbaum Associates.

Kirby, K.N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. Cognition, 51, 1-28.

Krauth, J., & Berchtold-Neumann, M. (1988). A model for disjunctive reasoning. Zeitschrift fur Psychologie, 196, 361-370.

McKenzie, C.R.M., Ferreira, V.S., Mikkelsen, L.A., McDermott, K.L., & Skrabbe, R.P. (2001). Do conditional statements target rare events? Organizational Behavior and Human Decision Processes, 85, 291-309.

McKenzie, C.R.M., & Mikkelsen, L.A. (2000). The psychological side of Hempel's paradox of confirmation. Psychonomic Bulletin and Review, 7, 360-366.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Oaksford, M. (2002a). Contrast classes and matching bias as explanations of the effects of negation on conditional reasoning. Thinking and Reasoning, 8, 13-151.

Oaksford, M. (2002b). Predicting the results of reasoning experiments: Reply to Feeney and Handley (2000). Quarterly Journal of Experimental Psychology, *55A*, 793-798.

Oaksford, M. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. Psychological Review, *101*, 608-631.

Oaksford, M., & Chater, N. (1995). Information gain explains relevance which explains the selection task. Cognition, *57*, 97-108.

Oaksford, M. & Chater, N. (1996). Rational explanation of the selection task. Psychological Review, *103*, 381-391.

Oaksford, M. & Chater, N. (2003). Optimal data selection: Revision, review and re-evaluation. Psychonomic Bulletin and Review, *10*, 289-318.

Oaksford, M., Chater, N., & Grainger, R. (1999). Probabilistic effects in data selection. Thinking and Reasoning, *5*, 193-243.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). Journal of Experimental Psychology: Learning, Memory, and Cognition, *23*, 441-458.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. Journal of Experimental Psychology: Learning, Memory, and Cognition, *18*, 835-854.

Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. Memory and Cognition, *31*, 143-154.

Over, D.E., & Green, D.W. (2001). Contingency, causation, and adaptive inference. Psychological Review, *108*, 682-684.

Roberts, M.J. (1998a). How should relevance be defined? What does inspection time measure? A reply to Evans. Quarterly Journal of Experimental Psychology, 51A, 815-817.

Roberts, M.J. (1998b). Inspection times and the selection task: Are they relevant? Quarterly Journal of Experimental Psychology, 51A, 781-810.

Roberts, M.J. (2002). The elusive matching bias effect in the disjunctive selection task. Experimental Psychology, 49, 89-97.

Roberts, M.J., & Newton, E.J. (2001). Inspection times, the change task, and the rapid response selection task. Quarterly Journal of Experimental Psychology, 54A, 1031-1048.

Stanovich, K.E., & West, R.F. (1998). Cognitive ability and variation in selection task performance. Thinking and Reasoning, 4, 193-230.

Stenning, K., & van Lambalgen, M. (2002). Semantics as a foundation for psychology: A case study of Wason's selection task. Journal of Logic, Language and Information, 10, 273-317.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), New horizons in psychology, Vol. I. Harmondsworth: Penguin.

Wason, P.C., & Evans, J.St.B.T. (1975). Dual processes in reasoning? Cognition, 3, 141-154.

Wason, P.C., & Johnson-Laird, P.N. (1972). Psychology of reasoning: Structure and content. London: Batsford.

Yama, H. (2001). Matching versus optimal data selection in the Wason selection task. Thinking and Reasoning, 7, 295-311.

Footnotes

¹ We note that current controversy surrounding the generality of matching bias to disjunctive selection tasks within the full negations paradigm (e.g., Roberts, 2002) has a critical bearing on Evans' claim that the basis of matching is that the topic of a sentence is the same irrespective of whether or not it has been negated. More evidence is clearly needed to resolve this controversy and to determine whether or not the H-A theory's claims regarding the origin of matching need to be qualified.

² Oaksford and Chater's (e.g., 1994, 1995) analysis also entails a rarity assumption, which is that most properties of the world (including the properties described by p and q in selection task studies) apply to a small set of objects, and that people's strategies for testing or framing hypotheses are, by default, adapted to situations where rarity holds (for supporting evidence see Anderson & Sheu, 1995; McKenzie, Ferreira, Mikkelsen, McDermott, & Skrabale, 2001; McKenzie & Mikkelsen, 2000).

³ In Experiment 1 a constant of 0.4 was used for both the facing-side and the hidden-side data transformations. In Experiment 2 a constant of 0.6 was used for the facing-side transformation, and one of 0.2 was used for the hidden-side transformation.

⁴ One interesting observation in relation to the Experiment 1 results presented in Table 3 (and indeed the Experiment 2 results in Table 6) is that it seems as if more references are made to the facing and hidden sides of selected cards than rejected ones when such cards should, in fact, be rejected according to the H-A model. We note, however, that Evans (e.g., 1996, 1998a) has clarified that "relevance" effects can extend beyond cards that are cued by the matching and the if heuristics. For example, a key claim of Evans (1996) is that even when selection patterns are seen to vary considerably across different thematic contents, selections can still be interpreted as arising from relevance judgements. In addition, we note that it is possible that a subset of individuals

in our experiments were pursuing full-blown and effortful logical analyses of cards (cf. Stanovich & West, 1998). The presence of such individuals in “selected” cells corresponding to infrequently selected cards could provide an alternative account of why more references are made to the facing and hidden sides of such cards by a minority of selectors relative to a majority of rejectors.

⁵ We are grateful to Mike Oaksford for pointing out to us this discrepancy between our protocol-based findings and the evidence forwarded by Feeney and Handley (2000).

Table 1

(a) The Four Types of Conditional Rule Employed in Experiments 1 and 2, Including the Standard Terminology that is Used to Describe Associated Cards, and (b) the Method Used for Assessing Matching Bias on Antecedent and Consequent Cases in Experiments 1 and 2

(a)	Logical Case			
	TA	FA	TC	FC
1. If A then 3	A (p)	J (not-p)	3 (q)	7 (not-q)
2. If E then not 5	E (p)	L (not-p)	2 (not-q)	5 (q)
3. If not S then 9	D (not-p)	S (p)	9 (q)	4 (not-q)
4. If not N then not 8	T (not-p)	N (p)	1 (not-q)	8 (q)

(b)

One-tailed Wilcoxon signed-rank comparisons for:

(i) Matching antecedent (TA:1 + TA:2 + FA:3 + FA:4) versus mismatching antecedent (TA:3 + TA:4 + FA:1 + FA:2), and

(ii) Matching consequent (TC:1 + TC:3 + FC:2 + FC:4) versus mismatching consequent (TC:2 + TC:4 + FC:1 + FC:3)

Note. TA: true antecedent; FA: false antecedent; TC: true consequent; FC: false consequent. Bold type indicates the elements on cards that match those in the presented rule. The logical combination of cards to select for each rule is TA and FC.

Table 2

Percentage Frequencies of Card Selections and Overall Mean References to Facing and to Hidden Sides for Experiment 1, by Items (N = 30)

Rule		TA Card		FA Card		TC Card		FC Card	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. If p then q	Frequency of Selections	90.0		16.7		60.0		10.0	
	Facing Side - ND	1.37	0.85	0.57	0.94	0.90	0.71	0.53	0.90
	Facing Side - TD	1.19		0.29		0.67		0.28	
	Hidden Side - ND	0.87	0.78	0.17	0.46	0.53	0.51	0.17	0.38
	Hidden Side - TD	0.67		0.08		0.38		0.09	
2. If p then not q	Frequency of Selections	86.7		10.0		10.0		63.3	
	Facing Side - ND	1.53	1.14	0.37	0.67	0.37	0.67	0.80	0.85
	Facing Side - TD	1.22		0.19		0.19		0.53	
	Hidden Side - ND	1.13	0.86	0.07	0.25	0.10	0.31	0.53	0.63
	Hidden Side - TD	0.89		0.04		0.06		0.34	
3. If not p then q	Frequency of Selections	60.0		43.3		66.7		43.3	
	Facing Side - ND	1.30	1.29	1.07	1.08	1.17	1.26	0.90	1.27
	Facing Side - TD	0.86		0.70		0.75		0.49	
	Hidden Side - ND	0.87	1.07	0.83	1.21	0.73	1.02	0.47	0.86
	Hidden Side - TD	0.56		0.43		0.39		0.23	
4. If not p then not q	Frequency of Selections	60.0		46.7		33.3		53.3	
	Facing Side - ND	1.13	1.14	1.10	0.92	0.70	0.84	1.10	1.03
	Facing Side - TD	0.75		0.78		0.43		0.75	
	Hidden Side - ND	0.47	0.86	0.60	0.72	0.20	0.41	0.63	0.72
	Hidden Side - TD	0.25		0.38		0.11		0.39	
Mean		74.2		29.2		42.5		42.5	

Note. ND = natural data; TD = transformed data (in original units).

Table 3

Mean References to Facing and to Hidden Sides (by Items) for Selected and Non-Selected Cards, for Experiment 1

Rule	Card	Selected					Non-selected				
		N	Facing		Hidden		N	Facing		Hidden	
			ND	TD	ND	TD		ND	TD	ND	TD
1. If p then q	TA	27	1.44	1.30	0.93	0.72	3	0.67	0.51	0.33	0.20
	FA	5	1.60	1.11	0.80	0.53	25	0.36	0.19	0.04	0.02
	TC	18	1.33	1.26	0.78	0.67	12	0.25	0.15	1.67	0.09
	FC	3	2.00	1.64	0.67	0.51	27	0.37	0.20	0.11	0.06
2. If p then not q	TA	26	1.54	1.34	1.12	0.89	4	1.50	0.75	1.25	0.89
	FA	3	2.00	2.00	0.67	0.51	27	0.19	0.10	0.00	-0.01
	TC	3	2.00	2.00	1.00	1.01	27	0.19	0.10	0.00	-0.01
	FC	19	1.05	0.89	0.74	0.58	11	0.36	0.15	0.18	0.07
3. If not p then q	TA	18	2.00	1.74	1.33	1.08	12	0.25	0.15	1.67	0.09
	FA	13	1.54	1.42	1.31	0.89	17	0.71	0.32	0.47	0.20
	TC	20	1.40	1.01	0.90	0.53	10	0.70	0.36	0.40	0.18
	FC	13	1.62	1.15	1.08	0.75	17	0.35	0.18	0.00	-0.01
4. If not p then not q	TA	18	1.50	1.05	0.72	0.41	12	0.58	0.38	0.08	0.05
	FA	14	1.50	1.38	1.00	0.80	16	0.75	0.41	0.25	0.13
	TC	10	1.20	0.92	0.40	0.26	20	0.45	0.26	0.10	0.06
	FC	16	1.69	1.51	1.06	0.86	14	0.43	0.25	0.14	0.08

Note. ND = natural data; TD = transformed data (in original units).

Table 4

Mean References to Facing and to Hidden Sides (by Participants) for Selected Versus Non-Selected Cards, for Experiment 1

	Selected				Non-selected			
	Facing		Hidden		Facing		Hidden	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ND:	1.52	0.51	0.98	0.49	0.46	0.49	0.16	0.26
TD (log):	0.22	0.13	0.05	0.19	-0.19	0.22	-0.32	0.12
TD	1.26		0.73		0.24		0.07	

Note. ND = natural data; TD (log) = transformed data (in \log_{10} units); TD = transformed data (in original units).

Table 5

Percentage Frequencies of Card Selections and Overall Mean References to Facing and to Hidden Sides for Experiment 2, by Items ($N = 30$)

Rule		TA Card		FA Card		TC Card		FC Card	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. If p then q	Frequency of Selections	80.0		23.3		80.0		16.7	
	Facing Side - ND	1.57	0.82	0.93	0.74	1.43	0.77	0.90	0.71
	Facing Side - TD	1.44		0.75		1.26		0.72	
	Hidden Side - ND	0.67	0.55	0.17	0.38	0.60	0.50	0.10	0.31
	Hidden Side - TD	0.43		0.07		0.39		0.04	
2. If p then not q	Frequency of Selections	86.7		33.3		33.3		70.0	
	Facing Side - ND	1.87	0.90	0.83	0.70	0.93	0.83	1.37	0.81
	Facing Side - TD	1.74		0.66		0.75		1.22	
	Hidden Side - ND	1.10	0.96	0.27	0.45	0.30	0.65	0.67	0.61
	Hidden Side - TD	0.67		0.12		0.12		0.40	
3. If not p then q	Frequency of Selections	56.7		46.7		66.7		43.3	
	Facing Side - ND	1.43	1.33	1.30	0.95	1.40	1.10	0.77	0.57
	Facing Side - TD	1.18		1.06		1.14		0.63	
	Hidden Side - ND	0.90	1.24	0.53	0.68	0.67	0.92	0.37	0.56
	Hidden Side - TD	0.45		0.26		0.33		0.17	
4. If not p then not q	Frequency of Selections	56.7		50.0		40.0		50.0	
	Facing Side - ND	1.17	0.75	1.27	1.05	1.10	0.76	1.00	0.95
	Facing Side - TD	1.02		0.99		0.95		0.72	
	Hidden Side - ND	0.50	0.57	0.60	0.62	0.33	0.55	0.53	0.68
	Hidden Side - TD	0.27		0.34		0.15		0.26	
Mean		70.0		38.3		55.0		45.0	

Note. ND = natural data.; TD = transformed data (in original units).

Table 6

Mean References to Facing and to Hidden Sides (by Items) for Selected and Non-Selected Cards, for Experiment 2

Rule	Card	Selected					Non-selected				
		N	Facing		Hidden		N	Facing		Hidden	
			ND	TD	ND	TD		ND	TD	ND	TD
1. If p then q	TA	24	1.58	1.49	0.75	0.52	6	1.50	1.31	0.33	0.16
	FA	7	1.14	0.99	0.29	0.13	23	0.87	0.69	0.13	0.05
	TC	24	1.50	1.40	0.71	0.51	6	1.17	0.95	0.17	0.07
	FC	5	1.40	1.35	0.20	0.09	25	0.80	0.63	0.08	0.03
2. If p then not q	TA	26	1.81	1.64	1.12	0.67	4	2.25	2.09	1.00	0.69
	FA	10	1.20	1.10	0.50	0.29	20	0.65	0.47	0.15	0.06
	TC	10	1.30	1.10	0.80	0.45	20	0.75	0.60	0.05	0.02
	FC	21	1.48	1.31	0.81	0.45	9	1.11	0.99	0.33	0.16
3. If not p then q	TA	17	1.29	1.22	0.76	0.48	13	1.62	1.10	1.07	0.39
	FA	14	1.36	1.14	0.71	0.40	16	1.25	0.95	0.38	0.16
	TC	20	1.35	1.22	0.60	0.39	10	1.50	0.99	0.80	0.23
	FC	13	0.77	0.63	0.39	0.20	17	0.77	0.63	0.35	0.15
4. If not p then not q	TA	17	1.17	1.31	0.71	0.46	13	0.85	0.69	0.23	0.10
	FA	15	1.53	1.35	0.87	0.61	15	1.00	0.66	0.33	0.16
	TC	12	1.17	1.14	0.58	0.31	18	1.06	0.81	0.17	0.07
	FC	15	0.93	0.72	0.67	0.36	15	1.07	0.72	0.40	0.18

Note. ND = natural data.; TD = transformed data (in original units).

Table 7

Mean References to Facing and to Hidden Sides (by Participants) for Selected Versus Non-Selected Cards, for Experiment 2

	Selected				Non-selected			
	Facing		Hidden		Facing		Hidden	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ND:	1.33	0.47	0.67	0.47	1.10	0.48	0.37	0.27
TD (log):	0.23	0.12	-0.27	0.25	0.17	0.14	-0.43	0.20
TD	1.08		0.34		0.88		0.17	

Note. ND = natural data; TD (log) = transformed data (in \log_{10} units); TD = transformed data (in original units).

Table 8

Mean References to Facing and to Hidden Sides (by Participants) for Cards Represented in Mental Models Versus Cards Not Represented in Mental Models, for (a) Experiment 1, and (b) Experiment 2

(a)	Represented in Mental Models				Not Represented in Mental Models			
	Facing		Hidden		Facing		Hidden	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ND:	1.04	0.56	0.63	0.45	0.59	0.59	0.22	0.35
TD (log):	0.04	0.18	-0.11	0.18	-0.15	0.23	-0.30	0.15
TD	0.70		0.38		0.30		0.10	

(b)	Represented in Mental Models				Not Represented in Mental Models			
	Facing		Hidden		Facing		Hidden	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ND:	1.32	0.51	0.62	0.36	0.86	0.52	0.23	0.27
TD (log):	0.23	0.12	-0.28	0.21	0.11	0.17	-0.53	0.20
TD	1.08		0.30		0.68		0.10	

Note. ND = natural data; TD (log) = transformed data (in \log_{10} units); TD = transformed data (in original units).